

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-039343

(43)Date of publication of application : 12.02.1999

(51)Int.Cl.

G06F 17/30
H04N 5/76

(21)Application number : 09-209735

(71)Applicant : MEDIA RINKU SYST:KK

(22)Date of filing : 17.07.1997

(72)Inventor : TSUCHIYA HARUKI

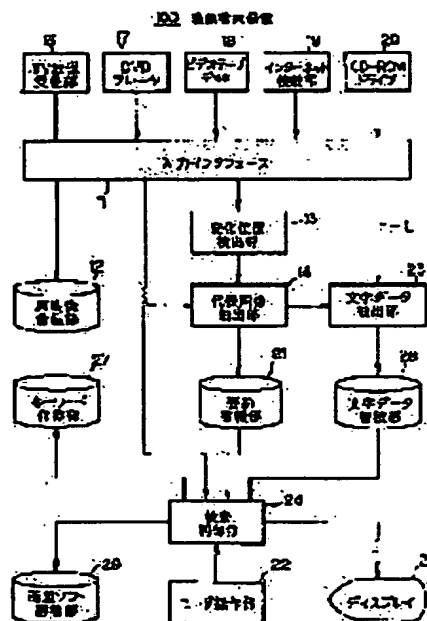
(54) VIDEO RETRIEVAL DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a suitable video retrieval device by using it for speedily detecting a video which a viewer desires from a large quantity of spread videos.

SOLUTION: A change position detection part 13 detects the switch of a scene. A representative picture extraction part 14 extracts one to several pieces of representative picture VS(video software) from the respective scenes. A character data extraction part 23 extracts character data from them. Character data is accumulated in a character data accumulation part 23 with information on the extracted position. When a user desires to check something in business or personally, accumulated character data is retrieved with a keyword showing it.

The representative picture VD which is hit and original video software VS are displayed on a display 26. Data can be selected from the display of a table.



LEGAL STATUS

[Date of request for examination] 15.07.2004

[Date of sending the examiner's decision of rejection] 13.02.2007

[Kind of final disposal of application other than

the examiner's decision of rejection or
application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Image retrieval equipment characterized by having an alphabetic character data extraction means to extract alphabetic data from image software, an alphabetic character data accumulation means to accumulate said alphabetic data with the extract positional information, a selection means to choose a desired thing out of said accumulated alphabetic data, and the cash-drawer means that pulls out the image concerning the this chosen alphabetic data out of said image software.

[Claim 2] Image retrieval equipment according to claim 1 with which said alphabetic character data extraction means is characterized by extracting said alphabetic data from the representation image extracted from said image software.

[Claim 3] Image retrieval equipment according to claim 1 or 2 characterized by said cash-drawer means pulling out the representation image concerning said selected alphabetic data.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention is used for finding quickly out that which the viewer wants out of the image circulated in large quantities in detail about image retrieval equipment, and relates to suitable image retrieval equipment. In addition, on these specifications, the thing equivalent to one television broadcasting is called a "image" or "screen", and what put two or more [of these] in a row is called "image." What added the sound signal, the caption code (title sign), etc. to this is called "image software (drawing 3 VS (this sign VS omits hereafter))." These are supplied with a motion-picture film, a video tape, television broadcasting, DVD (a digital videodisc or digital bar SATAIZU disk), the Internet, a computer storage, etc.

[0002]

[Description of the Prior Art] A movie, a sport relay broadcast, etc. are the time when image software is supplied in large quantities. Otherwise, the image software of very various fields, such as a drama, music, culture, and a report, exists, and it is expected that the amount of supply will increase further from now on. The utility value of these images software is very high as it is also in the proverb of seeing is believing. make it the thing of work -- make it individual hobby level -- the value will be very high, if these are grasped exactly and it can use as a library as the occasional decision ingredient.

[0003] Image software meets the passage of time amount. People need time amount, in order to see this. When the amount of supply becomes huge in this way, the image transcription etc. is carried out to the video tape like before, and it becomes impossible so, to say that it sees slowly later. Then, the keyword, the index for retrieval, etc. are attached to a required image (frame) and a required image, for example, and how to search later using this can be considered.

[0004]

[Problem(s) to be Solved by the Invention] However, this approach requires time amount and time and effort for the input of a keyword, an index, etc. Although the contents are checked about each image software and a keyword, an index, etc. are specifically attached for these whole or each image of every, aside from the ancient times which had little amount of supply, such an art is used in the future which is called multimedia age and whose amount increases today or further, and does not have *****, either. It always becomes a technical problem an index [what kind of keyword and] are attached.

[0005] Along with improvement in the calculation speed of a computer, it came to pull out what "a full-text search (full-text search) full-text search" Carries out the alphabetic data text itself, and corresponds by the information retrieval centering on an alphabetic character, without carrying out registration of indexation or a keyword beforehand to alphabetic data. In this case, if the object of the noun in the text is carried out, it will be known that quite exact retrieval can be performed, and if a user inputs the language currently used every day, the text containing that language will be searched at high speed.

[0006] It is possible like this that image retrieval also carries out "all image retrieval (full image search)." However, it is the processing that whose a match is looked for about an image itself time amount requires considerably, and it is common for it to be meaningless. For this reason, a certain

option is required.

[0007] The purpose of this invention is to use it for finding out quickly that which the viewer wants, and offer suitable image retrieval equipment out of such an image circulated in large quantities.

[0008]

[Means for Solving the Problem] By invention of claim 1, an alphabetic character data extraction means to extract alphabetic data from image software, an alphabetic character data accumulation means to accumulate said alphabetic data with the information on the extract location, a selection means to choose a desired thing out of said accumulated alphabetic data, and the cash-drawer means that pulls out the image concerning the this chosen alphabetic data out of said image software are used for the above-mentioned purpose achievement.

[0009] Moreover, in invention of claim 2, said alphabetic character data extraction means extracts said alphabetic data from the representation image extracted from said image software. Moreover, in invention of claim 3, the cash-drawer means said to claim 1 or claim 2 pulls out the representation image concerning said selected alphabetic data.

[0010] This invention is observing the alphabetic character contained in image software. Here, the alphabetic character in image software has the following descriptions.

(1) It is used for an alphabetic character telling important information, such as a title and a performer, on the movie.

(2) The title of a foreign movie is made so that conversation and a story may be told in the minimum alphabetic character.

(3) The alphabetic character is used for arranging the outline of an incident by the news of TV (television). The alphabetic character is used here in order to adjust information in a compact.

(4) As for the result of a horse race etc., numerical information is displayed on a screen in many cases as a result of a sport.

[0011] Thus, the alphabetic character in it is used for transfer of information with explicit semantics to the image software itself having told delicate feeling and sensuous things. By this invention, paying attention to this alphabetic data, this alphabetic data is extracted and accumulated by the configuration given in a claim, and facilities are given to a viewer by so to speak considering as the database of an image and an image.

[0012]

[Embodiment of the Invention] Hereafter, it explains based on the example of a gestalt of illustration implementation of the detail of this invention. The outline of introduction and the example of a gestalt of operation (drawing 1 image retrieval equipment 100) is explained.

(1) Original image software, such as received TV broadcast (television broadcasting), is supplied and accumulated in the original image are recording section 12 through the input interface 11 (although the original image are recording section 12 should consider as the "Hara image software are recording section" correctly, since it was redundancy, it omitted "software"). moreover, the change location detecting element 13 is also supplied -- having -- a scene (a scene, cut) -- changing -- etc. -- the location where the condition of image software changes is detected. Similarly the image software supplied by another gestalt, for example, a video tape, DVD, the Internet, CD-ROM, etc. is incorporated from the input interface 11.

[0013] (2) If it is detected that the change of state, for example, a scene, changed by the change location detecting element 13, the representation image extract section 14 will extract the representation image VD (drawing 2 (following and sign VD abbreviation)) following this. A representation image is the static image of one sheet or the short dynamic image for about 2 - 5 seconds taken out from near [the] a change-of-state location, and the set is accumulated here as an epitome VE (drawing 2 (following sign VE abbreviation)). An epitome becomes the about 1/ten to 1/100 amount of data compared with original image software. Therefore, next processing becomes simple. It is used also as a decision ingredient of whether to see original image software.

[0014] - A representation image may be extracted from original image software a fixed period. the alphabetic character in image software is comparatively long -- it is indicated by time amount. This is

because time amount until a viewer finishes reading the displayed text is taken into consideration, and if it says conversely, the same alphabetic character is displayed in the meantime. therefore, if it is alike and a still picture or an animation is extracted at the rate of one sheet, this time interval, for example, 5 seconds, 10 seconds, and the alphabetic character contained in that image software can carry out a ***** extract. When based on this technique, it is not necessary to extract a representation image, and an image is extracted with a fixed time interval from a original image, and the alphabetic character contained in it is extracted. In addition, it may be accumulated and the image extracted here may also be used as an epitome.

[0015] (3) Extract an alphabetic character from a representation image. Character recognition software is used and alphabetic data is picked out from the extracted representation image. The taken-out alphabetic data is accumulated as a keyword. It is also good to limit this keyword only to a noun depending on the case. In case a keyword is begun and said epitome and original image software are accumulated, the channel number, the broadcast date, the time second, the frame number, etc. attach the header information which can maintain mutual link relation.

[0016] The accumulated keyword, an epitome, and original image software are used as the following.

(1) Drive in as a keyword the language which expresses the event directly, for example, the language like the free keyword of PATOLIS, at every need about the theme against which the viewer bears an interest by the individual thing, concerning work to know whether related image software is accumulated.

(2) It collates with the alphabetic data with which this was accumulated. A viewer sets up the Ruhr of retrieval of full coincidence and partial coincidence and others. If there are some corresponding, next the epitome corresponding to it will be taken out and it will check the contents of the image.

(3) An epitome is seen, if it is thought that it is a thing about the theme for which it asks, the original image software will be taken out and the contents will be checked.

[0017] An epitome is the magnitude of 1/10 to 1/100 of a original image as above-mentioned, and seeing this does not take the time amount like **. Therefore, there is little time and effort which requires for retrieval the direction to which made the two-step style in this way, and it was presupposed that it discerns whether it is that for which it asks in an epitome first of all, and it ends.

[0018] When there are some which register a keyword beforehand and correspond to this, it is also good to record independently the table-of-contents information about the image software etc. For example, an one day viewer checks an are recording condition for this. The appropriate thing is displayed on a screen by list, a viewer is made to choose and the epitome and original image software are displayed. A keyword is not inputted, but the alphabetic data accumulated is displayed by the order of the Japanese syllabary, and a viewer chooses a desired thing out of it, and may be made to display the epitome and original image software.

[0019] Hereafter, the detail of the image retrieval equipment 100 of the example of a gestalt of operation is quoted, and drawing is explained. That is, the image software supplied from TV broadcast receive section 16, the DVD player 17, the video tape deck 18, the Internet connectivity section 19, and CD-ROM drive 20 and others is incorporated by image retrieval equipment 100 through the input interface 11 in drawing 1.

[0020] TV broadcast receive section 16 has many receiving channels. For example, terrestrial TV broadcast (VHF, UHF), satellite broadcasting service, a cable TV, etc. receive the analog and digital TV broadcast for which a viewer asks about two or more channels to two or more coincidence. From the Internet connectivity section 19, the image software on the Internet for which it asks is periodically incorporated following actuation of a viewer.

[0021] The image software incorporated from these is stored in the original image are recording section 12. Are recording is good to carry out by digitizing. The thing of an analog form is digitized with the input interface 11, and is accumulated in the original image are recording section 12. In order to stop the amount of data, it is good to carry out a data compression by the method of MUSE, MPEG, JPEG, and others. The image software on digital television broadcast and the Internet is good at the **. Image software is accumulated in the original image are recording section 12 as a break and every one file for

example, per 10 minutes. Each receiving date, the receiving time of the initial data of a file, a channel number, etc. record table-of-contents information (index) required for next read-out on a file allocation table (FAT).

[0022] Storing in the original image are recording section 12 is continuously carried out to some extent like 72 hours about the specified channel. Speech information is also recorded. In addition, a viewer sets up the date which records, time of day, and the channel for record by the user control unit 22. There are some time zones which turn out that it is not necessary to record from the start, and there are also some programs in which a viewer does not get interested. Then, not only at time but at the time, if it is made even a part unit is fine and possible [a time zone setup] for it, the next contents check activity of record will become easy. A time zone configuration procedure is the same as the existing video tape recorder etc., and good.

[0023] record (are recording) -- being cyclic (cyclic) -- it is good to carry out. For example, record of the memory capacity of the original image are recording section 12 will be enabled for five-channel and continuation three days. When there is time amount, a viewer checks the alphabetic data accumulated, an epitome, and original image software, and eliminates a unnecessary thing. The storage region of the original image are recording section 12 and the epitome are recording section 21 is vacant now. Storage capacity may be lost by a certain thing [that carry out period continuation and the contents of broadcast are recorded]. in this case, the inside of a recorded image -- most -- being old -- although -- a top -- the following new video signal -- overwrite -- it carries out.

[0024] What is necessary is for there to be no need of specifying the date time of day to record, and to specify only the channel (channel which wants to supervise) for which it asks by the user control unit 22, if it carries out like this. But there is a time zone which does not need to be recorded from the start as mentioned above. Therefore, even when recording cyclically, there is little storage capacity which the direction whose setup of the time zone of record needlessness was enabled consumes, and it ends.

[0025] Return and the change location detecting element 13 will notify that to drawing 1 at the representation image extract section 14, if a change of state arises to the attention element with which it was specified in the television broadcasting then incorporated or the DVD receipt image software currently then reproduced.

[0026] This invention extracts the alphabetic data contained in image software, accumulates this, and utilizes it as a key of image retrieval. Alphabetic data may be soon extracted from the original image software, as mentioned above, but (drawing 1 broken line L) as mentioned above, in order to make the check of the contents of record easy, and in order to simplify processing, as shown in drawing 2 , extracts the representation image VD from original image software, and extracts alphabetic data out of this representation image VD here (the sign VD of a representation image etc. is omitted suitably below).

[0027] Detection of a change-of-state location is for giving the motive of this representation image extract, and the representation image extract section 14 extracts a representation image following this. Moreover, not only the still picture of one sheet but the animation for about several seconds concerning the change location shall be contained in each representation image.

[0028] Detectable various elements (description) are contained in image software (program). For example, the element in which the sound of people's voice, music, a cheer, and pulse nature and much other detection are possible exists in voice (sound). Moreover, some detectable elements -- that the brightness of a screen changes suddenly and the color of the whole screen changes [standing it still or] suddenly -- exist in an image. A representation image extract here is performed in order to take out the alphabetic data which exists in image software. Therefore, it is good to choose the element which the change produces in relation to the appearance of the new alphabetic character to a screen also about the element to which its attention is paid.

[0029] For example, after in the case of a news program reading of the news manuscript by announcer and the axle-pin rake is completed and voice breaks off, a screen projects a new alphabetic character in the form which the following news manuscript read out and synchronized with initiation. Therefore, change of an element called intermittence of this voice can also be used for the motive of a

representation image extract. The inclination for an educational program etc. to be the same is seen.

[0030] Change of an element and the relevance of an alphabetic character appearance change according to the genre (field) of image software. So, if the property of the image software for are recording at that time is clear from a race card etc., in case said channel for record is set up, the element to which its attention is paid is made into which, or it carries out as [set / a viewer].

[0031] There is also an element independent of a property or a field. When saying that the object of record is not specified, i.e., it records continuously about some channels etc. and the alphabetic data in it is taken out as a search key, it is good to pay one's attention to such a universal element. The example is shown below.

(1) Cut down simply the image at the time of logging Hara image software initiation of a head image as a title (representation image). When the broadcast start time of the program concerned can be grasped by a race card etc., the head image is extracted. For example, if it divides per 10 minutes and prolonged continuation record is carried out, the head image of each break will be extracted as a title.

[0032] (2) Cut down an image periodically for every fixed time amount from periodic logging image software. If the epitome which consists of the representation image of M sheets is generated about the original image software which consists of images of N sheets, the image in every N/M sheets will be taken out from original image software as a representation image. This is the simplest approach. In addition, by the picture compression technique, such as MPEG, JPEG, and MUSE, a perfect image is transmitted to how many sheet cage, and difference with a front image is transmitted in the meantime. Therefore, about such image data, the perfect image sent to several of the sheets one sheet may be extracted as a representation image.

[0033] When taking out alphabetic data from image software soon, the technique of this periodic logging is useful without generating an epitome. That is, although the alphabetic character in image software is displayed since it reads to a viewer and is surely given, it requires appropriate time amount for people reading a thing. For this reason, the alphabetic character of a screen is also displayed in consideration of it comparatively long time amount, for example, 5 seconds, 10 seconds, and often. Therefore, if in the case of television for 5 seconds then an image will be taken out and inspected at the rate of one sheet to 300 frames supposing one sheet is displayed for 10 seconds every 150 frames, ***** of an alphabetic character which appears in the image software can be extracted.

[0034] So, when based on this technique, an epitome is not generated but the direct alphabetic character extract from original image software is performed. But the epitome is also the decision ingredient of whether to see original image software as above-mentioned. Therefore, alphabetic data is considered as the direct extract from the original image said here, the element which will generate the epitome which, on the other hand, expresses the contents of the image software concerned exactly reasonable is chosen independently, the change of state is made into a motive, and you may make it generate the epitome of the image software.

[0035] In this case, there is the following as an example of the element to which its attention is paid.

Genre of an image Attention element news Cut drama with a pattern (flip) Cut with a title Cut documentary with voice Cut English conversation with voice Cut sport with a title Applause, the cut which a cheer arises, and its circumference (voice Climax)

Animation Cut with a title Cut which is standing it still for a long time Cut TV shopping with voice Cut with a title (information, such as a price, is in sight)

Song program Cut from which music begins (from voice to distinction)

Educational program Cut variety show with a pattern Cut orchestra which a cheer arises Place which music starts (periodic extract)

Weather report Stationary cut (pattern (flip) = plate which drew an alphabetic character and drawing.) A speaker etc. uses it by TV.

In addition, an attention element, a predetermined location, etc. change with properties of a program and a broadcast channel in this way. Then, it is good to perform epitome generation and a check several times and to decide a desirable attention element, a predetermined location, etc. in it.

[0036] (3) When the logging scene for every scene changes, say in many cases that it projects an

alphabetic character new in an image. Therefore, it is also good to detect this, to take out the representation image of number of sheets suitably for every scene, and to investigate the existence of alphabetic data. Although it is arbitrary from which location in a scene a representation image is taken, supposing a scene changes by CS1 and CS2 of drawing 3, for example, the center PC of a scene etc. will become one candidate with them.

[0037] Distinction from which a scene changes is performed as the following.

(1) The intensive processing image of a pixel has two-dimensional breadth needless to say. For example, NTSC system is the assembly of a pixel (a pixel, point) which is about called 250 dot x525 line, there are lightness, saturation, and a hue in each pixel, and 30 images which consist of a set which is such a pixel are transmitted in 1 second.

[0038] Needless to say, there are many pixels. Then, concentration-ization of a pixel is performed here for improvement in the speed of processing, and simplification. For example, if image data will become 1/16 of amounts and same processing will be carried out about the pixel of 8x8 supposing it calculates an average value in the sum of the pixel of every direction 4x4, image data will become 1/64.

[0039] This concentration-ization abstracts a loose change of an image and has the operation which makes easy distinction from which a scene changes. That is, a scene changes, distinction measures fundamentally the pixel and pixel of two images which get mixed up, and it discerns whether a big difference is there. If the data of both pixel are compared in some fields of an image and many are specifically common, I will think that it is an image only in one scene which the person moved, without a background changing there.

[0040] however, for example, ** zoom-in / down, and ** -- when there is modification of the sense of a camera made slowly and the pixel comparison of the image of order is carried out without carrying out intensive processing here, even when appearance is slight -- pixel level -- the part of ***** -- data -- being different -- actual -- a scene -- on the way -- it comes out, and although it is, a scene will change and these many will be judged accidentally [be / it / a location].

[0041] If the same thing is looked at on the concentration--ization-processed coarse pixel level, since the pixel is coarse, the image to a certain number of sheets has the same data, and is changeless. So, these are correctly judged to be an image belonging to the same scene. Then, they are the value of these coarse pixels, and a (tx, y) as pretreatment first.

It takes out here. It is t:time of day here. x y: Coordinate of the image after intensive processing a: Value of the color of the point (x y). a -- R, G, and B -- the -- or it puts in ** picking -- C1, R+C2, G+C3, and B -- then, good (as for R, G, and B, the value of three-primary-colors information, and C1, C2 and C3 are weighting factors).

[0042] (2) Ask for the difference between each of the pixel and each pixel of the image before that for every image about the image data which performed concentration-ized processing of the direction of time amount, next the above-mentioned concentration-ized processing. And it judges whether it is magnitude to the extent that you may judge that a scene changes and the difference is a location. If you want to simplify processing, the threshold fixed about the size of the difference at this time is defined. The change rate of a scene judges with the image having changed there about a certain number in the coarse pixel which constitutes an image, when the difference was over the threshold.

[0043] But even if it belongs to the same scene, the magnitude of the difference between pixels changes with the contents. If 1 coma 1 coma is the scene which changes early, the difference of the image of order is large, and difference is small if late. Therefore, although processing simple like the above is also law on the other hand, here adds the following processing further in order to extract a representation image here more exactly.

the time difference of "image -- comparative processing" -- data of each point which constitutes the image of one sheet It asks for time difference about a (tx, y). Namely, $d(tx, y) = a(tx, y) - a(t - ** tx, y)$ *****. However, **t: It is proper time amount width of face. This shows the difference between pixels of the image of two sheets with which only **t adjoins each other in time (variation).

[0044] It asks for this difference d (tx, y) about a predetermined period, for example, each image for 1 minute. It will be [Equation 1] if expressed with a formula.

$$D a(t) = \sum_{x,y} d(t,x,y)$$

It becomes.

[0045] Drawing 3 is the example of the difference $D a$ about the concentration-ized image for 1 minute (t) for which carried out in this way and it asked. The location CS1 and CS2 where a value is large, i.e., the locations where the difference of an image is large, means that an image has a certain big change there, and possibility that the scene has changed is high there. Therefore, between these CS1 and CS2 is presumed to be one scene, and the representation image of number of sheets is suitably extracted out of this. the time difference of "image which these say to the above-mentioned header -- it is comparison processing."

[0046] In addition, it asks for the difference accompanying a change of a screen from this, and you may make it detect the change rate of a scene about each frame which constitutes image software paying attention to the video signal of horizontal scanning line 1 duty for example, near an image center. The video signal of this horizontal scanning line 1 duty is divided into the section of N individual, and, specifically, the sum or the average is calculated about each section. And it asks for the difference to the average of the section of the screen before that concerned for every section. It is good, though this difference is totaled for every frame, it asks for the location equivalent to CS2 of the location 1 where that value is large, i.e., CS of drawing 3, and it is the change location of the same scene as the above of this.

[0047] Return and each representation image extracted by doing in this way are supplied to drawing 1 at the epitome are recording section 21 and the alphabetic character data extraction section 23. The epitome are recording section 21 holds the set of the supplied representation image as an epitome corresponding to the image software concerned. In addition, the time of the recording start of the taken-out date and a original image is recorded on a file so that the relation between each accumulated representation image and its original image may be known. Moreover, the information on the ejection location of the representation image is given to each representation image. This can be expressed with the number of progress seconds from there, a frame number, etc. on the basis of the recording start time of for example, a original image.

[0048] From the supplied representation image, the alphabetic character data extraction section 23 detects original image software to direct alphabetic data. There are some gestalten in the alphabetic data contained in image software. Therefore, detection here is also performed according to this gestalt. Since alphabetic data may be contained in one image software with two or more gestalten, it is good to also perform detection by two or more technique.

[0049] (1) When the alphabetic character is incorporated into the image as an image, i.e., an image or dot data.

In this case, an alphabetic character is detected with the application of the technique of Kanji OCR (optical alphabetic character read). Generally an alphabetic character is brought near and displayed on the lower part of an image, or right and left. Therefore, as long as precision may become coarse somewhat, the existence of an alphabetic character may be inspected about these parts. Then, simplification of processing and improvement in the speed can be performed.

[0050] (2) When a character code is added.

Like a teletext title, alphabetic data (title data) may be supplied by the character code as information on the skimmer of an image into the frame signal at the time of perpendicular feedback. In this case, it is easy. What is necessary is just to accumulate these character codes in the alphabetic character data accumulation section 28 in order with the positional information of an image. Furthermore, if it is made to develop, about the image software with which this character code is added, a representation image may not be extracted and you may extract only paying attention to the positional information of a character code and an image.

[0051] Return and the user control unit 22 become drawing 1 from a keyboard, a mouse, etc. again. Following these actuation by the viewer, the retrieval control section 24 performs the ejection of the

ejection of the alphabetic data accumulated in the alphabetic character data accumulation section 28, the epitome corresponding to selected alphabetic data, and original image software, a display on a display 26, etc. Moreover, 27 is a keyword attaching part and the keyword beforehand inputted by the viewer is held here. The retrieval control section 24 also performs registration processing of this keyword, and collating processing with these and the are recording data of the alphabetic character data accumulation section 28.

[0052] Like the above-mentioned, indexes, such as a broadcast date, a time second, and a frame number, are given to each, and a original image and the alphabetic data summarized and detected are accumulated in it at the original image are recording section 12, the epitome are recording section 21, or the alphabetic character data accumulation section 28. These are the databases of image software itself. Therefore, the method of the use is various. Although explanation of the above-mentioned rough actuation is overlapped, it will be as follows if an example is raised.

[0053] (A) Whenever a viewer is the need, input and search a keyword.

This search method is the most general database retrieval technique. If the theme which is an individual thing and a viewer wants for something to investigate specifically comes out, concerning (1) work, the theme will be driven in from the user control unit 22 supposing some language which will be expressed directly.

[0054] (2) The retrieval control section 24 collates this keyword with the alphabetic data accumulated in the alphabetic character data accumulation section 28 following this, and inspects the existence of applicable data. Collating is carried out with the gestalt for which full coincidence and partial coincidence, others, and a viewer ask.

(3) The retrieval control section 24 will indicate the table-of-contents information by list at a display 26, if the corresponding epitome and image software exist. If it held in media, such as DVD, the volume name showing the DVD etc., a title name, and a track number and others are indicated by list at a display 26.

[0055] (4) A viewer chooses a desired thing out of this list. If broadcasting hours, a DVD title, etc. make a key table-of-contents information currently taken out to the chart, the approach to the epitome for which it asks will become easy. The retrieval control section 24 takes out the epitome corresponding to the alphabetic data which the viewer chose from the epitome are recording section 21, and displays it on a display 26.

[0056] Various methods of presentation of an epitome are also assumed. For example, the same with turning over a page, whenever the viewer pressed the key etc., every one representation image may be displayed. A representation image may be displayed on an automatic turning-over type one after another with a certain amount of speed. Needless to say, there is story nature in image software. Therefore, even if it displays one after another at the early rate, a viewer can understand the contents of the epitome exactly unexpectedly. Functions, such as slow playback and a review, are also prepared and it is also good to enable it to use this if needed.

[0057] (5) See a proper number of epitomes, or see to the last, and the image software or its part directs the thing relevant to the theme for which it asked, or playback of the original image software from the location which took out the representation image when it could judge that it was likely to be related. The epitome has magnitude of 1/100 from 1/10 of original image software, and much time amount is not taken to see. So, the direction made into two steps in this way can shorten the time of concentration to the original image for which it asks.

[0058] Following directions of the above-mentioned viewer, the retrieval control section 24 reads the original image software after the epitome ejection location from the original image are recording section 12 one by one, and displays it on a display 26. This processing is performed like regeneration with a video tape deck, others, and the usual visual equipment. Here, it is user-friendly when the key of a rapid traverse, a review, and a pause and others is prepared.

[0059] (B) It is a list display about what registers the keyword beforehand and corresponded.

In this case, the keyword showing the matter which a viewer wants to watch is previously accumulated in the keyword attaching part 27. For example, if one day and a viewer do confirmation operation, a list

indication of the table-of-contents information on the epitome containing the alphabetic data which is in agreement with the keyword registered, or original image software will be given at a display 26 following this.

[0060] That is, in our working, it needs individually or the informational class to watch is restricted to some extent. Moreover, it is also rather troublesome to repeat and input the keyword same every day. Then, the keyword used for this seed fixed target is beforehand registered into the keyword attaching part 27 in this way. And what corresponds when a viewer wants is indicated by list at a display 26. For example, if you want to always grasp the image software about a patent and it will want, keywords, such as "a patent", a "industrial right", and the "Patent Office", will be registered into the keyword attaching part 27, and corresponding image software will be ****(ed). If the identifier of a politician, a famous man, and a film star is registered, it can be begun immediately to find the image which he wants to know.

[0061] (C) It is a list display about the alphabetic data accumulated.

For example, the alphabetic data accumulated is displayed on a display 26 in the order of the Japanese syllabary. A viewer chooses a desired thing out of it, and may be made to display the epitome and original image software corresponding to it.

[0062] If a representation image to save, an epitome, image software, etc. come out after checking an epitome and image software by these various technique, the key non-illustrated "preservation" etc. will be operated. The retrieval control section 24 stores them in the request software are recording section 29 following this. This becomes behind, and when required, it can refer to at any time.

[0063] In addition, the audio playback approach is arbitrary. For example, what is necessary is just to reproduce at the same rate, when reproducing original image software. However, when reproducing an epitome, playback is impossible in the original **. Then, the voice data for several seconds near [each] a representation image extract location is taken out, for example, and it brings [a little] forward whether it is the same as the time of a original image in this, and reproduces. Moreover, after digitizing, it culls out suitably, and it may accumulate and it may be reproduced.

[0064] A modification is described. Epitome playback may display four representation images and six on one screen of a display 26 collectively. The representation image has story nature as above-mentioned. Therefore, even if two or more these are displayed on a display 26 at once, a viewer can realize the contents easily. If it takes that there is a story into consideration, it can say that to do in this way is more intelligible.

[0065] If it followed, for example, X representation images were displayed on the screen at once, the epitome concerned is simply reproducible by the time amount of 1 for X at the time of every one sheet. Since a duration is short and ends, display time may be lengthened about the display screen which compounded this representation image of X sheets to one sheet. Thereby, an understanding of the image software concerned becomes easy further.

[0066] You may realize as equipment of dedication and this invention may be realized as an application program of a personal computer. An exclusive chip is manufactured, and it may include in a personal computer, or may include in a DVD player, a video tape deck, a game player, etc., or you may carry out.

[0067] About one image software, each representation image which makes change of a different element a motive may be extracted, and you may accumulate as a separate epitome. In the case of the contents check of record, each epitome can be reproduced and a subsequent check can be performed in the direction which is easy to understand.

[0068] The magnetic tape of a video tape and others for home use may be used for the original image are recording section 12 and the epitome are recording section 21. When chart lasting time is insufficient, two or more decks (drive) are arranged. Are recording of original image software takes appropriate storage capacity. Therefore, it is thought that the above-mentioned cyclic method of the record is good. However, about an epitome and a character code, since necessary storage capacity is small, you may accumulate considerably for a long period of time.

[0069] Although a check becomes a little difficult since original image software disappears by

overwrite, it can be grasped to some extent what the contents of the image software with which leaving alphabetic data and an epitome also corresponded to the keyword were. They can also be asked to a key to a broadcasting agency and data supply origin. Therefore, even if it makes it this operation format, a use is large as a database.

[0070] It is better not to accumulate the thing without the need, since it grows also into the cause which increases a retrieval noise also about an epitome and a character code. For this reason, it is good also for processing in which will accumulate only what was decided for a viewer to inspect a part for what day, or the accumulated thing, and to save it, and the remainder will be eliminated on that spot (retrieval noise = data which have separated from the retrieval purpose although it corresponds to the keyword concerned).

[0071]

[Effect of the Invention] As explained above, the alphabetic character which exists in image software took out and accumulated these paying attention to the property to express the contents of the image software directly, and made this invention available as a database. Therefore, when there is need always, a viewer is carrying out the keyword of a desired retrieval format, for example, the language currently used every day, etc., can pull out a request easily out of the image software supplied in large quantities, and can utilize these for work at an individual application.

[Translation done.]

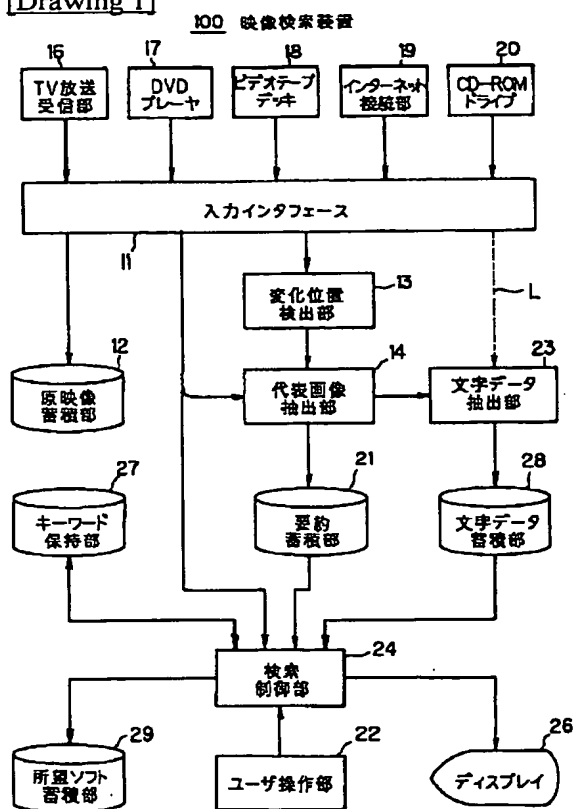
* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

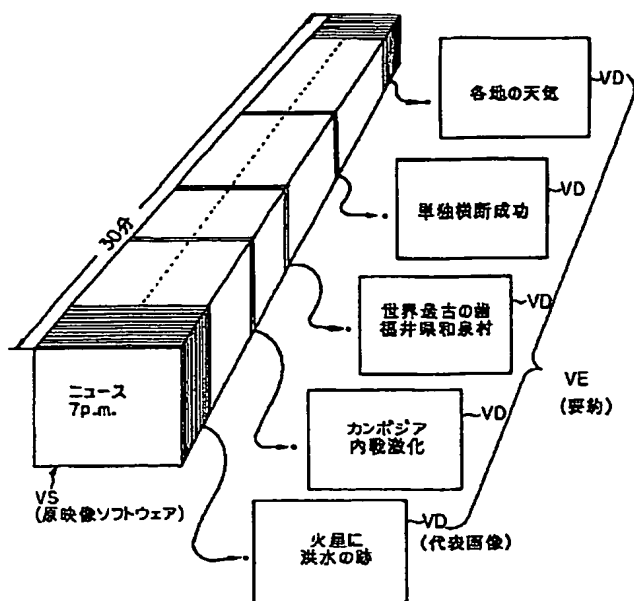
1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DRAWINGS

[Drawing 1]

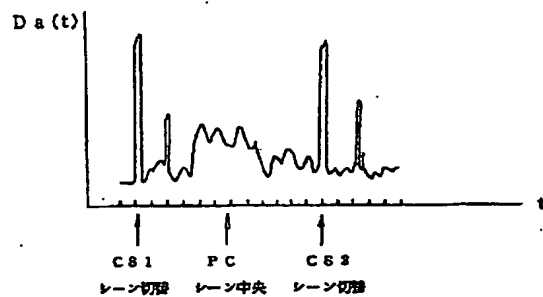


[Drawing 2]



[Drawing 3]

$D_a(t)$... 時刻 t と 時刻 $t + \Delta t$ の 映像の差分



[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-39343

(43) 公開日 平成11年(1999) 2月12日

(51) Int. Cl. ⁴	識別記号	P I	
G 0 6 F 17/30		G 0 6 F 15/40	3 7 0 D
H 0 4 N 5/76		H 0 4 N 5/76	B

審査請求 未請求 請求項の数 3 F D (全 9 頁)

(21) 出願番号 特願平9-208735

(22) 出願日 平成9年(1997) 7月17日

(71) 出願人 591015854

株式会社メディア・リンク・システム
東京都中央区東日本橋2-2-10 東日本
橋オリモビル

(72) 発明者 松屋 治紀

東京都中央区東日本橋2-2-10 株式会
社メディア・リンク・システム内

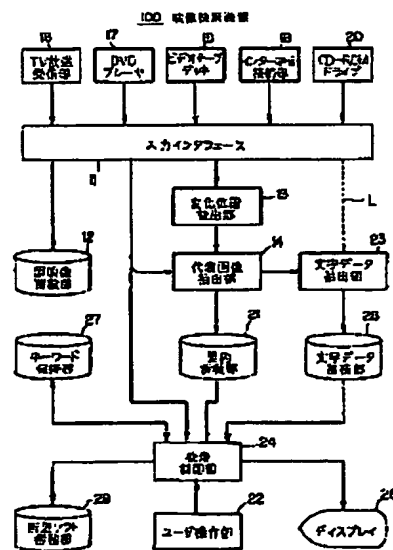
(74) 代理人 弁理士 松井 晃一

(54) 【発明の名称】 映像検索装置

(57) 【要約】

【課題】 大量に流布される映像の中から、視聴者が欲しているものを素早く見つけ出すのに使用して好適な映像検索装置を提供。

【解決手段】 シーンの切り替わり等を変化位置検出部13で検出。代表画像抽出部14で夫々のシーン等から1枚〜数枚の代表画像V Sを抽出。文字データ抽出部23で夫々から文字データ抽出。その抽出位置の情報と共にこれを文字データ蓄積部23に蓄積。仕事中或いは個人的に何か調べたいことがあったら、それを表わすキーワードで蓄積された文字データ検索。ヒットしたその代表画像V D、原映像ソフトウェアV Sをディスプレイ26に表示。一覧表示からの選択等も可。



(2)

特開平11-39343

1

【特許請求の範囲】

【請求項1】 映像ソフトウェアから文字データを抽出する文字データ抽出手段と、その抽出位置情報と共に前記文字データを蓄積する文字データ蓄積手段と、前記蓄積された文字データの中から所望のものを選択する選択手段と、該選択された文字データに係る画像を前記映像ソフトウェアの中から引出す引出し手段とを備えたことを特徴とする映像検索装置。

【請求項2】 前記文字データ抽出手段が、前記映像ソフトウェアから抽出した代表画像から前記文字データを抽出することを特徴とする請求項1に記載の映像検索装置。

【請求項3】 前記引出し手段が、前記選択された文字データに係る代表画像を引出すことを特徴とする請求項1又は請求項2に記載の映像検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は映像検索装置に関し、詳しくは、大量に流布される映像の中から視聴者が欲しているものを素早く見つけ出すのに使用して好適な映像検索装置に関する。なお本明細書では、テレビジョン放送1フレーム分に相当するものを「画像」又は「画面」と言い、これを複数枚連ねたものを「映像」と言う。これに音声信号、キャプションコード（字幕符号）等を付加したものを「映像ソフトウェア（図3VS（以下、この符号VSは省略））」と言う。これらは、映画フィルム、ビデオテープ、テレビジョン放送、DVD（デジタルビデオディスク）或いはデジタルバーサタイルディスク、インターネット、コンピュータ記憶媒体などで供給される。

【0002】

【従来の技術】映画やスポーツ中継など、映像ソフトウェアが大量に供給される時代になって来た。ほかに、ドラマ、音楽、教養、報道など、実に多岐多様な分野の映像ソフトウェアが存在し、その供給量は今後更に増大すると予想される。百聞は一見に如かずの諺にもあり、これら映像ソフトウェアの利用価値は非常に高い。仕事のことし、個人の趣味レベルにしろ、これらを的確に把握し、そのときどきの判断材料として、或いはライブラリとして利用出来るなら、その価値は非常に高い。

【0003】映像ソフトウェアは時間の流れに沿うものである。人はこれを見るために時間を必要とする。それ故、このように供給量が膨大になって来ると、従来のようにビデオテープに録画等しておいて、後でゆっくり見るという訳には行かなくなる。そこで、例えば必要な画像（フレーム）や映像にキーワード、検索用インデックス等を付けておき、これを使って後で検索を行うという手法が考えられる。

【0004】

2

【発明が解決しようとする課題】しかしこの方法は、キーワード、インデックス等の入力に時間と手間がかかる。具体的には、一つ一つの映像ソフトウェアについて内容を確認し、それら全体又は個々の画像毎にキーワード、インデックス等を付すのであるが、このような処理方法は、供給量が少なかった昔は兎も角、マルチメディア時代と言われる今日、或いは更に量が増える将来に於ては用い得べくもない。どんなキーワード、インデックスを付すかも常に課題になる。

【0005】コンピュータの計算速度の向上につれ、文字を中心にした情報検索では、文字データに対し予めインデックス化やキーワードの登録をせずに、文字データ本文そのものを「全文検索（フルテキストサーチ）」して該当するものを引出すようになった。この場合、本文中の単語を対象すればかなりの的確な検索が出来ることが知られており、利用者が日常使用している言葉を入力すると、その言葉を含んでいるテキストが高速で検索される。

【0006】これと同様に、画像検索でも「全画像検索（フルイメージサーチ）」をすることが考えられる。しかし、画像に関しては一致するものを探すこと自体がかなり時間の掛かる処理であり、またそれが無意味であることが多い。この為、何らかの別の方法が必要である。

【0007】本発明の目的は、このような大量に流布される映像の中から、視聴者が欲しているものを素早く見つけ出すのに使用して好適な映像検索装置を提供することにある。

【0008】

【課題を解決するための手段】上記目的達成のため請求項1の発明では、映像ソフトウェアから文字データを抽出する文字データ抽出手段と、その抽出位置の情報と共に前記文字データを蓄積する文字データ蓄積手段と、前記蓄積された文字データの中から所望のものを選択する選択手段と、該選択された文字データに係る画像を前記映像ソフトウェアの中から引出す引出し手段とを用いる。

【0009】また請求項2の発明では、前記文字データ抽出手段が、前記映像ソフトウェアから抽出した代表画像から前記文字データを抽出する。また請求項3の発明では、請求項1または請求項2にいう引出し手段が、前記選択された文字データに係る代表画像を引出す。

【0010】本発明は、映像ソフトウェアに含まれる文字に注目している。ここで、映像ソフトウェアのなかの文字は次のような特徴をもっている。

(1) 映画などでは、文字はタイトル、出演者などの重要な情報を伝えるのに利用されている。

(2) 外国映画の字幕は、最小限の文字で会話やストーリーを伝えるように作られている。

(3) TV（テレビジョン）のニュースなどでは、事件のあらましを要約するのに文字が利用されている。ここ

(3)

特開平11-39343

3

で文字はコンパクトに情報をとりまとめるために利用されている。

(4) スポーツの結果、競馬の結果などは、数値情報が画面に表示されることが多い。

【0011】このように、映像ソフトウェア自体は微妙な感情や感性的なことがらを伝えているのに対し、その中の文字は、明示的な意味のある情報の伝達に利用されている。本発明では、この文字データに着目し、請求項記載の構成によって、この文字データを抽出し、蓄積し、言わば画像、映像のデータベースとすることで、視聴者の利便を図る。

【0012】

【発明の実施の形態】以下、本発明の詳細を図示実施の形態例に基いて説明する。始めに、実施の形態例(図1映像検索装置100)の概要を説明する。

(1) 受信されたTV放送(テレビジョン放送)等の原映像ソフトウェアは、入力インタフェース11を介して原映像蓄積部12に供給され、蓄積される(原映像蓄積部12は、正しくは「原映像ソフトウェア蓄積部」とすべきだが冗長なので「ソフトウェア」を省略した)。また変化位置検出部13にも供給され、シーン(場面、カット)の切り替わりなど、映像ソフトウェアの状態が変化する位置が検出される。別の形態、例えばビデオテープ、DVD、インターネット、CD-ROMなどで供給された映像ソフトウェアも、同じく入力インタフェース11から取り込まれる。

【0013】(2) 変化位置検出部13によって状態変化、例えばシーンが切り替わったことが検出されると、これに反応して代表画像抽出部14が代表画像VD(図2(以下、符号VD省略))を抽出する。ここに代表画像とは、その状態変化位置付近から取り出される1枚の静止画像または2〜5秒程度の短い動画であり、その集合が要約VE(図2(以下符号VE省略))として蓄積される。要約は原映像ソフトウェアに比べ10分の1〜100分の1程度のデータ量になる。従ってこの後の処理が簡便になる。原映像ソフトウェアを見るか否かの判断材料としても利用される。

【0014】・代表画像は一定周期で原映像ソフトウェアから抽出しても良い。映像ソフトウェア中の文字は比較的時間表示される。これは、表示された文章を視聴者が読み終る迄の時間を考慮しているためで、逆に言えば、この間は同じ文字が表示されている。従ってこの時間間隔、例えば5秒とか、10秒とかに1枚の割合で、静止画か動画かを抽出すれば、その映像ソフトウェアに含まれる文字は殆んど漏らさず抽出できる。この手法によるときは代表画像を抽出する必要はなく、原映像から一定時間間隔で画像を抜き出して、その中に含まれる文字を抽出する。なおここで抽出した画像も、集積して要約として利用しても良い。

【0015】(3) 代表画像から文字を抽出する。文字

4

認識ソフトウェアを使用し、抽出した代表画像から文字データを取り出す。取り出した文字データはキーワードとして蓄積する。場合によってはこのキーワードを名詞だけに限定するのも良い。キーワードを始め、前記要約、原映像ソフトウェアを蓄積する際には、例えばチャネル番号、放送年月日、時分秒、フレーム番号など、相互のリンク関係を保てる見出し情報を付けておく。

【0016】蓄積されたキーワード、要約、原映像ソフトウェアは、例えば以下の如く利用する。

(1) 仕事に関し、或いは個人的なことで、視聴者が関心を持っているテーマについて、関連する映像ソフトウェアが蓄積されているかを知りたい場合、その事象を端的に表現する言葉、例えばPATOISのフリーキーワードの如き言葉を、必要の都度、キーワードとして打ち込む。

(2) これが蓄積された文字データに照合される。完全一致、部分一致その他の検索のルールは視聴者が設定する。該当するものがあつたら、次はそれに対応した要約を取り出し画像の内容を確認する。

(3) 要約を見て、それが所望するテーマに関するものだと考えられたら、その原映像ソフトウェアを取り出して内容を確認する。

【0017】前述のとおり要約は原映像の10分の1から100分の1の大きさで、これを見るにさほどの時間は掛からない。従ってこのように2段階にいて、先ずは要約で所望するものか否かを見定めることとした方が、検索に掛かる手間が少なくて済む。

【0018】予めキーワードを登録し、これに該当するものがあつたらその映像ソフトウェアに関する目次情報等を別に記録しておくのも良い。これを例えば一日一回視聴者が蓄積状態をチェックするようにする。該当したものを一覧で画面に表示し、視聴者に選択させ、その要約、原映像ソフトウェアが表示されるようにする。キーワードを入力するのではなく、蓄積されている文字データを例えば五十音順で表示し、視聴者が、その中から所望のものを選択して、その要約、原映像ソフトウェアを表示するようにしても良い。

【0019】以下、実施の形態例の映像検索装置100の詳細を図を引用して説明する。即ち図1に於て、TV放送受信部16、DVDプレーヤ17、ビデオテープデッキ18、インターネット接続部19、CD-ROMドライブ20その他から供給される映像ソフトウェアは、入力インタフェース11を介し映像検索装置100に取り込まれる。

【0020】TV放送受信部16は、多数の受信チャネルを備えている。例えば地上波TV放送(VHF、UHF)、衛星放送、ケーブルTVなど、複数のチャネルについて視聴者が所望するアナログ、デジタルのTV放送を複数同時に受信する。インターネット接続部19からは、定期的に、或いは視聴者の操作に応じて、所望す

(4)

特開平11-39343

5

るインターネット上の映像ソフトウェアが取り込まれる。

【0021】これらから取り込まれた映像ソフトウェアは原映像番組部12に蓄えられる。蓄積はデジタル化して行なうと良い。アナログ方式のものは入力インタフェース11でデジタル化して原映像番組部12に蓄積する。データ量を抑えるため、MUSE、MPEG、JPEGその他の方式でデータ圧縮しておくとも良い。デジタルテレビジョン放送、インターネット上の映像ソフトウェアはその中で良い。映像ソフトウェアは、例えば10分単位に区切り、一つ一つのファイルとして原映像番組部12に蓄積する。夫々の受信年月日、ファイルの先頭データの受信時刻、チャンネル番号など、後での読み出しに必要な目次情報（インデックス）を、ファイルアロケーションテーブル（FAT）に記録する。

【0022】原映像番組部12への格納は、指定されたチャンネルについて、例えば72時間というように、ある程度連続して実施する。音声情報も記録する。なお、記録をする年月日、時刻、記録対象チャンネルは、ユーザ操作部22により視聴者が設定する。始めから記録する必要がないと判っている時間帯もあるし、視聴者が関心を持たない番組もある。そこで、単に日時だけでなく、時、分単位くらいまで細かく時間帯設定可能にしておく、後の記録内容確認作業が楽になる。時間帯設定手順は既存のビデオテープレコーダー等と同じで良い。

【0023】記録（蓄積）はサイクリック（循環的）に行なうと良い。例えば原映像番組部12の記憶容量を、5チャンネル分、連続3日記録可能としておく。視聴者は時間があるとき、蓄積されている文字データ、要約、原映像ソフトウェアを確認し、不用のものを消去する。これで原映像番組部12と要約番組部21の記憶領域が空く。ある期間連続して放送内容が記録されることで記憶容量が狭くなることもある。この場合は、記録済み映像の中の一番古いものの上に、次の新しい映像信号を重ね書きする。

【0024】こうすれば、記録する年月日時刻等を指定する必要は無く、所望するチャンネル（監視をしておきたいチャンネル）のみユーザ操作部22で指定しておけば良い。尤も、前述したとおり始めから記録しなくてもよい時間帯がある。従ってサイクリックに記録する場合でも、記録不要の時間帯を設定可能にしておいた方が消費する記憶容量が少なくて済む。

【0025】図1に戻り、変化位置検出部13は、そのとき取り込まれているテレビジョン放送、或いはそのとき再生されているDVR記録映像ソフトウェアの中の、指定された注目要素に状態変化が生じたら、その旨を代表画像抽出部14に通知する。

【0026】本発明は映像ソフトウェアの中に含まれる文字データを抽出し、これを蓄積して映像検索のキーとして活用する。文字データは、前述したように、元の映

6

像ソフトウェアから直接抽出しても良いが（図1参照し）、ここでは、前述した如く記録内容のチェックを容易にするために、また処理を簡略化するために、図2に示すように、原映像ソフトウェアから代表画像VDを抽出し、この代表画像VDの中から文字データを抽出する（以下代表画像の符号VD等は適宜省略）。

【0027】状態変化位置の検出は、この代表画像抽出の動作を与えるためのもので、これに依拠して、代表画像抽出部14が代表画像を抽出する。また夫々の代表画像には、1枚の静止画だけでなく、その変化位置に係る数秒程度の動画も含まれるものとする。

【0028】映像ソフトウェア（放送番組）には検出可能な様々な要素（特徴）が含まれる。例えば音声（音）には、人の声、音楽、歓声、パルス性の音、その他多数の検出可能な要素が存在する。また映像には、静止していることや、画面の明るさが急変すること、画面全体の色が急変することなど、検出可能な要素が幾つか存在する。ここでの代表画像抽出は、映像ソフトウェアの中に存在する文字データを取り出すために行なうものである。従って、着目する要素についても、画面への新たな文字の出現に関連してその変化が生ずる要素を選みと良い。

【0029】例えばニュース番組の場合、アナウンサー、キャスターによるニュース原稿の朗読が終了して声が途切れた後に、次のニュース原稿の読み上げ開始に同期した形で、画面に新たな文字が映出される。従って、この音声の断続という要素の変化も、代表画像抽出の動機に利用出来る。教育番組なども同じ傾向が見られる。

【0030】要素の変化と文字出現の関連性は、映像ソフトウェアのジャンル（分野）に応じて変る。それ故、番組表などからそのときの蓄積対象映像ソフトウェアの特性が明らかなら、前記記録対象チャンネルを設定する際、着目する要素をどれにするか、視聴者が設定可能であるようにしておく。

【0031】特性や分野に依存しない要素もある。記録の対象を特定しない、即ち幾つかのチャンネル等について連続して記録をし、その中の文字データを検索キーとして取り出す、というようなときは、このような普遍的要素に着目すると良い。以下にその例を示す。

（1）先頭画像の切り出し

原映像ソフトウェア開始時の画像をタイトル（代表画像）として単独に切り出す。番組表などで当該番組の放送開始時刻が把握出来るときは、その先頭画像を抽出する。例えば10分単位で区切って長時間連続記録するなら、各区切りの先頭画像をタイトルとして抽出する。

【0032】（2）周期的切り出し

映像ソフトウェアから一定時間ごとに周期的に画像を切り出す。N枚の画像で構成される原映像ソフトウェアについて、M枚の代表画像から成る要約を生成するならば、原映像ソフトウェアからN/M枚ごとの画像を代表

(5)

特開平11-39343

7

8

画像としてとりだす。これは最も単純な方法である。なお、MPEG、JPEG、MUSEなどの画像圧縮手法では、何枚かごとに完全な画像が送信され、その間は、前の画像との差分が送信される。従って、このような画像データについては、その何枚かに1枚送られて来る完全な画像を代表画像として抽出しても良い。

【0033】この局所的切り出しという手法は、要約を生成しないで、映像ソフトウェアから直に文字データを取り出すとき有用である。即ち、映像ソフトウェアの中の文字は、まさしく視聴者に読んで貰う為に表示されるものであるが、人がものを読むにはそれなりの時間が掛かる。このため、画面の文字もそれを考慮して、割に長い時間、例えば5秒とか10秒とか表示されることが多い。従って、仮に5秒間とすれば、テレビジョンの場合で、150フレーム毎に1枚、10秒表示されるとすれば、

映像のジャンル	注目要素
ニュース	パターン（フリップ）のあるカット
ドラマ	字幕のあるカット 音声のあるカット
ドキュメンタリー	音声のあるカット
英会話	字幕のあるカット
スポーツ	拍手、歓声の上がるカットとその周辺 （音声クライマックス）
アニメ	字幕のあるカット 長く静止しているカット 音声のあるカット
TVショッピング	字幕のあるカット （価格などの情報が見える）
歌番組	音楽の始まるカット （音声から判別）
教育番組	パターンのあるカット
バラエティショー	歓声の上がるカット
オーケストラ	音楽がスタートするところ （局所的抽出）
天気予報	静止したカット

（パターン（フリップ）＝文字や図を書いた板。TVで話し手などが使用。）

なお、このように注目要素、所定位置等は番組、放送チャンネルの特性によって異なる。そこで要約生成、チェックを何回か行なって、その中で好ましい注目要素、所定位置等を決めるとよい。

【0036】（3）シーンごとの切り出し

シーンが変わったとき、画像中に新たな文字が映出されるということが多い。従って、これを検出し、シーン毎に適宜枚数の代表画像を取り出して文字データの有無を調べるのも良い。シーンの中のどの位置から代表画像を取るかは任意だが、例えば図3のCS1、CS2でシーンが変わるとしたら、それらと、シーン中央PCなどが一つの候補になる。

【0037】シーンの切り替わりの判別は、例えば以下の如く行なう。

*ば300フレームに1枚の割合で画像を取り出して検査すれば、その映像ソフトウェア中に現れる文字の殆んどを抽出できる。

【0034】それ故、この手法によるときは、要約を生成せず、原映像ソフトウェアからの直の文字抽出を実行する。尤も前述のとおり、要約は原映像ソフトウェアを見るべきか否かの判断材料にもなっている。従って、文字データを、ここにいう原映像からの直接抽出とし、その一方で、当該映像ソフトウェアの内容を尤も的確に表現する要約を生成するであろう要素を別を選んで、その状態変化を動機にしてその映像ソフトウェアの要約を生成するようにしても良い。

【0035】この場合、着目する要素の例としては以下のようなものがある。

（1）ピクセルの集約処理

映像は言うまでもなく2次元的な広がりを持つ。例えばNTSC方式は、凡そ250ドット×525ラインと云うようなピクセル（画素、点）の集まりであり、夫々のピクセルには明度、彩度、色相があり、このようなピクセルの集合からなる画像が1秒間に30枚送信される。

【0038】ピクセルの数は、言うまでもなく多い。そこで、処理の高速化、簡素化の為、ここではピクセルの集約化を行う。例えば縦横4×4のピクセルの和なり、平均値なりを求めるとすると、映像データは1/16の量になり、8×8のピクセルについて同様の処理をすると、映像データは1/64になる。

50 【0039】この集約化は、画像の緩やかな変化を抽象

9

し、シーンの切り替わりの判別を容易にする作用がある。即ち、シーンの切り替わり判別は、基本的には、前後する二つの画像のピクセルとピクセルとを比較し、そこに大きな差があるか否かを見極めるものである。具体的には、例えば画像の幾つかの領域で両者のピクセルのデータを比較し、多くが共通していたら、そこは、例えば背景が変わらずに人物が移動しただけとかの、一つのシーンの中の画像であると考ええる。

【0040】しかし、例えば

① ズームアップ/ダウンや

② ゆっくりしたカメラの向きの変更

があったとき、ここでいう集約処理をしないで前後の画像のピクセル比較をすると、見た目は僅かでも、ピクセルレベルでは殆どどの部分でデータが相違して、実際にはシーンの途中であるのに、これらの多くがシーンの切り替わり位置であると誤って判断されてしまう。

【0041】おなじものを、集約化処理した粗いピクセルレベルで見ると、ある枚数までの画像は、ピクセルが粗いためデータが同じで変化がない。それ故、これらは同じシーンに属する画像だと正しく判断される。そこで先ず前処理として、これら粗いピクセルの値、

$a(t, x, y)$

をここで取り出す。ここに、 t ：時刻 x, y ：集約処理後の画像の座標 a ：その点 (x, y) の色の値。 a は、 R, G, B をその値取り入れるか、 $C_1 \cdot R + C_2 \cdot G + C_3 \cdot B$ とすればよい(R, G, B は3原色情報の値、 C_1, C_2, C_3 は重み係数である)。

【0042】(2)時間方向の集約化処理

次に、上記集約化処理を行なった映像データについて、各画像毎にその夫々のピクセルとその前の画像の夫々のピクセルとの間の差分を求める。そしてその差分が、シーンの切り替わり位置であると判断してよいほどの大きいか否かを判断する。処理を単純にしたいなら、このときの差分の大きさについて一定のしきい値を定めておく、画像を構成する粗いピクセルの中のある数について、その差分がしきい値を超えていたら、そこで画像が変わった、即ちシーンの切り替わりがあったと判定する。

【0043】尤も、同じシーンに属していても、その内容によってピクセル間の差分の大きさは異なる。1コマ1コマが早く移り変わるシーンなら前後の画像の差分は大きく、遅いものなら差分が小さい。従って、上記の如く単純な処理も一方法だが、ここは、よりの確に代表画像を抽出するべく、更に下記処理を加える。

「映像の時間的な差分比較の処理」一枚の画像を構成する各点のデータ $a(t, x, y)$ について、時間的な差分を求める。即ち

$d(t, x, y) = a(t, x, y) - a(t - \Delta t, x, y)$

を求める。但し、 Δt ：適宜の時間幅である。これは、時間的に Δt だけ隔り合う2枚の画像のピクセル相互の

(6)

特開平11-39343

10

差分(変化量)を示している。

【0044】この差分 $d(t, x, y)$ を、所定期間、例えば1分間の間の夫々の画像に関して求める。式で表わせば、

【数1】

$$Da(t) = \sum_{x, y} d(t, x, y)$$

となる。

【0045】図3は、このようにして求めた例えば1分間の集約化映像についての差分 $Da(t)$ の例である。値が大きい位置、即ち映像の差分が大きい位置CS1、CS2が、そこで画像に何らかの大きな変化があることを表わしており、そこではシーンが切り替わっている可能性が高い。従って、このCS1とCS2の間を一つのシーンと推定し、この中から適宜枚数の代表画像を抽出する。これらが上記見出しにいう「映像の時間的な差分比較処理」である。

【0046】なお映像ソフトウェアを構成する各フレームについて、例えば画像中央付近の水平走査線1本分の映像信号に着目し、これに対して画面の移り変わりに伴う差分を求め、シーンの切り替わりを検出するようにしても良い。具体的には、例えば、この水平走査線1本分の映像信号を、N個の区間に分け、夫々の区間について和或いは平均値を求める。そして、各区間毎に、その前の画面の当該区間の平均値に対する差分を求める。この差分を各フレーム毎に総和し、その値が大きくなっている位置、即ち、図3のCS1、或いはCS2に当たる位置を求め、これを上記同様のシーンの切り替え位置であるとしても良い。

【0047】図1に戻り、このようにして抽出された各代表画像は、要約書部21と文字データ抽出部23に供給される。要約書部21は、供給された代表画像の集合を当該映像ソフトウェアに対応した要約として保持する。なお、蓄積した各代表画像とその原映像との関係が判るように、ファイルには取り出した年月日、原映像の記録開始の時分を記録する。また、個々の代表画像には、その代表画像の取り出し位置の情報を付しておく。これは、例えば原映像の記録開始時分を基点とし、そこからの経過秒数、フレーム番号などで表わせる。

【0048】文字データ抽出部23は供給された代表画像から、或いは原映像ソフトウェアから直接文字データを検出する。映像ソフトウェアに含まれる文字データには幾つかの形態がある。従ってここでの検出もこの形態に合わせて行なう。一つの映像ソフトウェアに複数の形態で文字データが含まれることもあるから、検出も複数の手法で行なうと良い。

【0049】(1)文字が画像として、即ちイメージ或いはドットデータとして画像中に組み込まれている場合、

この場合は、漢字OCR(光学的文字読取り)の手法を

11

適用して文字を検出する。文字は一般に画像の下方や左右に寄せて表示される。従って、精度が多少粗くても構わなければ、これらの部分について文字の有無を検査しても良い。そうすれば処理の簡素化、高速化が出来る。

【0050】(2) 文字コードが付加されて来る場合。文字放送字幕のように、垂直同期時のフレーム信号の中に、画像のスキマの情報として文字データ(字幕データ)が文字コードで供給される場合もある。この場合は簡単である。これら文字コードを映像の位置情報とともに順に文字データ蓄積部28に蓄積していけば良い。更に発展させるなら、この文字コードが付加されている映像ソフトウェアについては、代表画像の抽出をせず、文字コードと映像の位置情報だけに注目して抽出を行っても良い。

【0051】再び図1に戻り、ユーザ操作部22は、キーボード、マウスなどからなる。視聴者によるこれらの操作に応動し、検索制御部24は、文字データ蓄積部28に蓄積された文字データの取り出し、選択された文字データに対応した要約、原映像ソフトウェアの取り出し、ディスプレイ26への表示等を行なう。また27はキーワード保持部で、予め視聴者によって入力されたキーワードがここに保持される。検索制御部24は、このキーワードの登録処理、これらと文字データ蓄積部28の蓄積データとの照合処理も行なう。

【0052】原映像、要約、検出された文字データは、前述の如く、夫々に、放送年月日、時分秒、フレーム番号等のインデックスが付され、原映像蓄積部12、要約蓄積部21または文字データ蓄積部28に蓄積される。これらは映像ソフトウェアのデータベースそのものである。従ってその利用の仕方は多様である。前述の大まかな動作の説明と重複するが、一例を上げれば以下のとおりである。

【0053】(A) 視聴者が必要の都度、キーワードを入力して検索。この検索方法は、最も一般的なデータベース検索手法である。具体的には、

(1) 仕事に関し、或いは個人的なことで、視聴者が何か調べたいテーマが出て来たら、そのテーマを端的に表現するであろう言葉を幾つか想定し、ユーザ操作部22から打ち込む。

【0054】(2) 検索制御部24は、これに応動して、文字データ蓄積部28に蓄積されている文字データにこのキーワードを照合し、該当データの有無を検査する。照合は、完全一致、部分一致その他、視聴者の希望する形態で実施する。

(3) 検索制御部24は、該当する要約、映像ソフトウェアが存在したら、その目次情報をディスプレイ26に一覧表示する。DVDなどの媒体に収容されたものなら、そのDVD等を表わすボリューム名やタイトル名、

(7)

特開平11-39343

12

トラック番号その他をディスプレイ26に一覧表示する。

【0055】(4) 視聴者は、この一覧の中から所望のものを選択する。放送時間、DVDタイトルなど、一覧表に出されている目次情報を手軽にすれば、所望する要約へのアプローチが容易になる。検索制御部24は、視聴者が選択した文字データに対応した要約を要約蓄積部21から取り出し、ディスプレイ26に表示する。

【0056】要約の表示方法も種々想定される。例えばページをめくるのと同じように、視聴者がキー等を押した都度、一つづつ代表画像を表示していても良い。自動めくり式に、ある程度の速さで次々と代表画像を表示していても良い。映像ソフトウェアには言うまでもなくストーリー性がある。従って、早い速度で次々と表示していても、視聴者は意外に的確に要約の内容を理解できる。スロー再生、レビュー等の機能も設けておき、必要に応じてこれを利用出来るようにしておくのも良い。

【0057】(5) 適宜の数の要約を見て、或いは最後まで見て、その映像ソフトウェア或いはその一部分が、所望したテーマに関連したもの、或いは関連していそうだと判断できたら、その代表画像を取り出した位置からの原映像ソフトウェアの再生を指示する。要約は原映像ソフトウェアの10分の1から100分の1の大きさになっており、見るのに多くの時間を要しない。それ故、このように2段階にした方が、所望する原映像への到達時間を短縮できる。

【0058】検索制御部24は、上記視聴者の指示に応動して、原映像蓄積部12からその要約取り出し位置以降の原映像ソフトウェアを順次読み出し、ディスプレイ26に表示する。この処理は、ビデオテープデッキその他、通常の映像機器での再生処理と同様に行なわれる。ここでも、早送り、レビュー、ポーズその他のキーを設けておく使い勝手が良い。

【0059】(B) 予めキーワードを登録しておいて、該当したものを一覧表示。

この場合は、視聴者が見張っていたい事項を表わすキーワードを先にキーワード保持部27に蓄積しておく。例えば一日一回、視聴者が確認操作をすると、これに応動して、登録されているキーワードに一致する文字データを含んでいた要約、或いは原映像ソフトウェアの目次情報が、ディスプレイ26に一覧表示される。

【0060】即ち、我々が仕事上、或いは個人的に必要とする、或いは見張っておきたい情報の種類は、ある程度限られている。また、毎日同じキーワードを繰り返し入力するのもしんどい。そこでこのように、この種定期的に使用されるキーワードを、キーワード保持部27に予め登録しておく。そして視聴者が欲するとき該当するものをディスプレイ26に一覧表示する。例え

50

(8)

特開平11-39343

13

ば、特許に関する映像ソフトウェアを常に把握しておきたいと欲するなら、「特許」「工業所有権」「特許庁」といったキーワードをキーワード保持部27に登録しておき、該当する映像ソフトウェアを捕獲する。政治家、有名人、映画スターの名前を登録しておけば、自分の知りたい映像をすぐに見つけさせる。

【0061】(C) 蓄積されている文字データを一覧表示。

例えば蓄積されている文字データをディスプレイ26に五十音順に表示する。視聴者が、その中から所望のものを選択し、それに対応した要約や原映像ソフトウェアを表示するようにしても良い。

【0062】これら各種手法で要約や映像ソフトウェアを確認した後、若し保存したい代表画像、要約、映像ソフトウェアなどが出来たら、不図示「保存」のキー等を実行する。検索制御部24はこれに反応して、それらを所望ソフト蓄積部29に格納する。これにより後になって必要なときいつでも参照出来る。

【0063】なお音声の再生方法は任意である。例えば原映像ソフトウェアを再生するときは、同じ速度で再生すれば良い。しかし要約を再生する場合は、元の倍で再生は出来ない。そこで、例えば夫々の代表画像抽出位置付近の数秒間の音声データを取り出し、これを原映像のときと同じか少し早めて再生する。また、デジタル化したあと適宜間引きして蓄積し、それを再生しても良い。

【0064】変形例について述べる。要約再生は代表画像4つとか6つとかをディスプレイ26の一つの画面に羅列して表示しても良い。前述のとおり代表画像はストーリー性を持っている。従って、これらが複数一度にディスプレイ26に表示されても、視聴者は容易にその内容を把握出来る。ストーリーがあることを考慮すると、このようにした方が判りやすいとも言える。

【0065】従って、例えば代表画像X枚を一度に画面に表示したとすれば、単純には1枚づつのときのX分の一の時間で当該要約を再生することが出来る。所要時間が短くて済むから、このX枚の代表画像を1枚に合成した表示画面については、表示時間を長くして良い。これにより一層当該映像ソフトウェアの理解が容易になる。

【0066】本発明は専用の装置として実現してもよいし、パーソナルコンピュータのアプリケーションプログラムとして実現してもよい。専用チップを製作し、パソコンに組み込んだり、DVDプレーヤ、ビデオテープデッキ、ゲームプレーヤ等に組み込んだりしても良い。

【0067】一つの映像ソフトウェアについて、異なる要素の変化を助観とする夫々の代表画像を抽出し、別個の要約として蓄積しても良い。記録内容チェックの際、夫々の要約を再生して、理解しやすい方でその後のチェックを実行することが出来る。

【0068】原映像蓄積部12、要約蓄積部21は、家庭用のビデオテープその他の磁気テープを用いても良

14

い。記録時間が足りないときは、デッキ(ドライブ)を複数個配置する。原映像ソフトウェアの蓄積にはそれなりの記憶容量を要する。従ってその記録は前述の循環的方式が良いと思われる。しかし、要約と文字コードについては、所要記憶容量が小さいから、かなり長期間蓄積しても構わない。

【0069】重ね書きで原映像ソフトウェアが消えてしまうので確認がやや困難にはなるが、文字データと要約を残しておくだけでも、キーワードに該当した映像ソフトウェアの中身が何であったかは、ある程度把握出来る。それらを手掛かりに、放送元、データ供給元へ問い合わせをすることも出来る。従って、この実施形式にしたとしても、データベースとして効用は大きい。

【0070】要約と文字コードについても、検索ノイズを増やす原因にも成るから、必要のないものは蓄積しない方が良い。この為、何日分か蓄積されたものを視聴者が検査し、保存すると決めたものだけ蓄積して、残りをその場で消去してしまうという処理にも良い(検索ノイズ=当該キーワードに該当してはいるが、検索目的からは外れているデータ)。

【0071】

【発明の効果】以上説明したように本発明は、映像ソフトウェア中に存在する文字が、その映像ソフトウェアの内容を端的に表わすという性質に着目し、これらを取り出して蓄積し、データベースとして利用可能にした。従って、視聴者は、何時でも必要があるときに、所望の検索形式、例えば日常使用されている言葉をキーワードするなど、大量に供給される映像ソフトウェアの中から容易に所望のもの引出すことが出来、仕様に、個人的用途にこれらを活用することが出来る。

【図面の簡単な説明】

【図1】本発明の実施の形態の一例を示すブロック図。

【図2】映像ソフトウェア、代表画像、文字データの例を示す概念図。

【図3】時刻tの映像データと、時刻t+Δtの映像データの差分とシーンの切り替わりとの関係を示す波形状図。

【符号の説明】

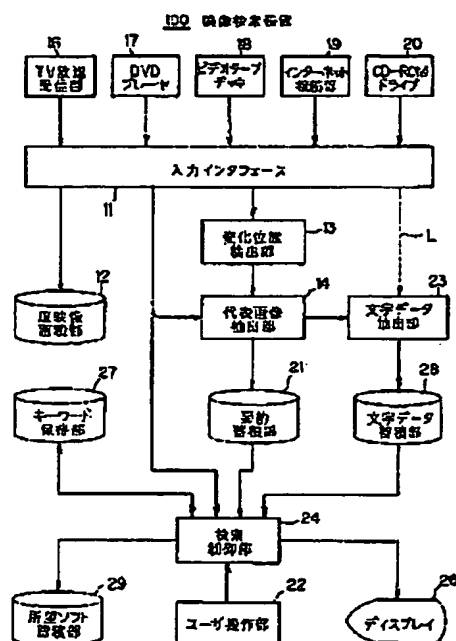
100…映像検索装置	11…入力インタフェース
12…原映像蓄積部	13…変化位置検出部
14…代表画像抽出部	16…TV放送受信部
17…DVDプレーヤ	18…ビデオテープデッキ
19…インターネット接続部	20…CD-ROMドライブ
21…要約蓄積部	22…ユーザ操作部

(9)

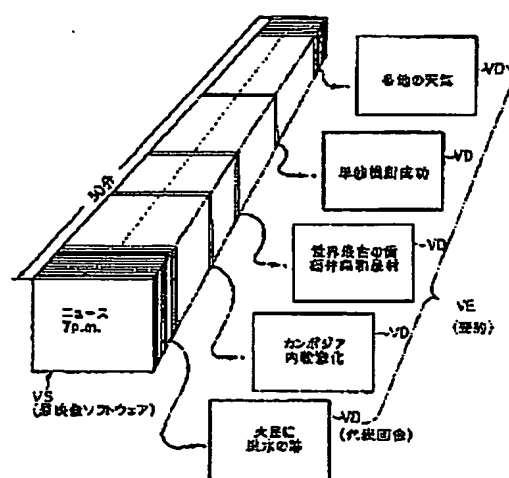
特開平11-39343

- 15
 23…文字データ抽出部
 26…ディスプレイ
 保持部
 28…文字データ蓄積部
 蓄積部
 CS…シーン切替
- 16
 24…検索制御部
 27…キーワード
 29…所望ソフト
 Da(t)…時刻t*
- *と時刻t+Δtの映像の差分
 PC…シーン中央
 L…文字データ直接抽出
 VE…要約
 ウェア
 VD…代表画像
 VS…映像ソフト

【図1】



【図2】



【図3】

Da(t) … 時刻tと時刻t+Δtの映像の差分

